

Submitted 2 November 2013

Topical Editor Decision: Reconsider after major revisions (further review by Editor and Referees) (30 Jan 2014) by Prof. Christopher Owen

Comments to the Author:

Dear Dr Svalgaard:

I have received two rather conflicting reviews from the referees of the manuscript you have under consideration for publication in *Annales Geophysicae*. One referee suggest only a very minor correction while the other recommends rejection after highlighting what s/he feels are a number of unacceptable aspects. Under normal circumstances, I would be minded to ask for a third referee report. You may wish to request I do that. However, it was rather difficult to find appropriate colleagues who were prepared to act as referee for this manuscript in the first place, so I am wary that this risks prolonging the assessment of this manuscript even further. I therefore propose that you look at and digest the critical comments of referee #1 and attempt to deal with these in a revised version. Of course you are welcome to provide a full rebuttal, in your accompanying response to the referee, to any points you feel are not valid, but it would be useful if we were able to move forward in a way that referee #1 felt able to withdraw his reservations.

Report #1

Submitted on 28 Dec 2013

Anonymous Referee #1

Anonymous during peer-review: Yes No

Anonymous in acknowledgements of published article: Yes No

Does the paper contain new data or new ideas or both of them? **Yes No**

Are these up to international standards? **Yes No**

Is the presentation clear? **Yes No**

Does the author reach substantial conclusions? **Yes No**

Is the length of the paper adequate? **Yes No**

Is the language fluent and precise? **Yes No**

Are the title and the abstract pertinent and understandable? **Yes No**

Is the size of each figure adequate to the quantity of data it contains? **Yes No**

Does the author give proper credit to related work and does he/she indicate clearly his/her own contribution? **Yes No**

Would you cite this paper as a scientific contribution? **May have potent after additional v and resubmission**

Recommendation to the Editor

The manuscript is not acceptable.

Comments to the Author

The author (Svalgaard) has provided evidence that certain duration of the Helsinki observatory magnetic data are improperly calibrated. The author's conclusion is based on expected correlations between different indices using data from other observatories. Implicit in such an analysis is the expectation that a correlation will be found. Of course, sometimes data are not correlated and when this happens for real physical reasons, then the data contain interesting information. I say this because I am very reluctant to agree that the author has *necessarily* identified a problem with the data. What he has done is make an inference. I might even agree with him on this inference. But calling the data "erroneous", as is done in the title of this manuscript and at many other points in the body of the manuscript is going too far. These parts of the manuscript need to be fixed, and the title needs to change to something more balanced.

Still, the author has certainly given some interesting evidence for a possible (even likely) problem. What I find puzzling is the personal nature of the manuscript. He directly urges Nevanlinna, who recently made public the data in question, to pursue this issue, see, for example, line 176 and line 314. Has the author (Svalgaard) been in communication with Nevanlinna on this subject?

I note that the author suggests that "metadata" be used to correct the miscalibration, line 180. The author says these metadata should be found and used to make the correction. There is no alternative, and the author needs to make this clear. Data that have problems cannot be "fixed" without actual auxiliary measurements, such as the calibration metadata discussed by Svalgaard. Anything else, such as estimates based on inference (like that reported in this manuscript) is not acceptable. Many things are not known. For example, are there multiple miscalibrations in the data? Is the calibration drifting around for some reason? When, exactly, did the miscalibration begin? When was it fixed? How do we know the answers to these questions? A statistical analysis (like that reported here) cannot answer these questions with precision. It can only help us understand that a possible problem exists. Unfortunately, as the manuscript is presently written, observatory workers might be tempted to "fix" their data archives. It has happened before, sometimes with unfortunate follow-on consequences. So, this issue must be clarified in this manuscript before it is considered acceptable for publication.

I do think, however, that a manuscript like this one can serve the role of alerting the geophysical community to a possible problem. With the compilation of similar reports for other historical data sets, researchers can better know how to proceed. But sometimes bad data simply remain bad. Yes, it is regrettable. Yes.

The changes that need to be made are significant. Until they are made, I recommend rejection.

Report #2

Submitted on 14 Jan 2014

Referee #2: Truls Lynne Hansen, truls.hansen@uit.no

Anonymous during peer-review: Yes No

Anonymous in acknowledgements of published article: Yes No

Does the paper contain new data or new ideas or both of them?	Yes No
Are these up to international standards?	Yes No
Is the presentation clear?	Yes No
Does the author reach substantial conclusions?	Yes No
Is the length of the paper adequate?	Yes No
Is the language fluent and precise?	Yes No
Are the title and the abstract pertinent and understandable?	Yes No
Is the size of each figure adequate to the quantity of data it contains?	Yes No
Does the author give proper credit to related work and does he/she indicate clearly his/her own contribution?	Yes No
Would you cite this paper as a scientific contribution?	Very important Fairly im

Recommendation to the Editor

The manuscript is acceptable with some corrections.

Comments to the Author

Congratulations! Though, could you please explain which sites are hidden behind the station codes PSM-VLJ-CLF in lines 148-149.

<<< The comments of referee \#1 are rather moot now as Lockwood et al. [Reconstruction of geomagnetic activity and near-Earth interplanetary conditions over the past 167 yr – Part 3] have already [in three papers with reference to my paper under review] accepted the conclusion of my paper as correct and have themselves [using independent comparisons with other stations and with the aa-index] corrected the Helsinki data very closely along the lines that I suggested. Since referee \#1 does not

have specific points that can be addressed [other than some philosophical remarks] I do not see how a meaningful rebuttal can be constructed as well of it not being necessary now that the goal of my paper has already been accomplished.
Correction of 'raw' data is often necessary and can be done with good results, see e.g. this recent article in your journal: <http://www.ann-geophys.net/32/19/2014/angeo-32-19-2014.pdf>

Topical Editor Decision: Reconsider after major revisions (further review by Editor and Referees) (02 Apr 2014) by Prof. Christopher Owen
Comments to the Author:
Dear Dr Svalgaard:

Thank you for your resubmission. I have passed your comments back to the critical referee, and, in addition, I have commissioned a third review of the paper and process to date. As you will see from the comments, there remains some feeling that you could do more to engage with the process and improve the manuscript. I feel therefore, in order to protect the integrity of the peer review process at the journal, that I need to ask you to make a more complete attempt to address the comments made to date by these 2 referees. I would appreciate your cooperation in this and look forward to the receipt of a revised manuscript and response.

Report #2

Submitted on 06 Mar 2014
Anonymous Referee #3

Anonymous during peer-review: Yes No

Anonymous in acknowledgements of published article: Yes No

Does the paper contain new data or new ideas or both of them?	Yes No
Are these up to international standards?	Yes No
Is the presentation clear?	Yes No
Does the author reach substantial conclusions?	Yes No
Is the length of the paper adequate?	Yes No
Is the language fluent and precise?	Yes No
Are the title and the abstract pertinent and understandable?	Yes No
Is the size of each figure adequate to the quantity of data it contains?	Yes No
Does the author give proper credit to related work and does he/she indicate clearly his/her own contribution?	Yes No

Would you cite this paper as a scientific contribution?

Fairly
Important

Fai

Recommendation to the Editor

The manuscript will be acceptable after minor revisions.

Comments to the Author

I have read the paper and the comments of referee 1 and 2. I will restrict my comments to referee 2's objection to the paper.

I understand what referee 2 is saying and he/she does make a very valid, and extremely important, point of principle. Recalibration of historic data is not something to be taken lightly and there are many past examples of where this has been done badly, sometimes based on the knowledge available at that time which turns out to be inadequate, and/or with certain assumptions and/or using limited statistics. The problem is that once a false correction has been made it is often remarkably difficult to track it down and reverse it. Thus corrections to data must be done carefully and openly with maximum rigour and with meta-data on how and why it was corrected kept and made readily available. For this reason I do agree with the referee that the language needs softening as the author should not be – nor want to be – as definite as he sounds in the present wording. For example “Recalibration of Magnetic Elements for Helsinki Data” would be a much more honest and acceptable title than “Errors in Scale Values for Magnetic Elements for Helsinki”. By asserting that there are definitely “errors” the author is stating that he is 100% certain that there is not another factor, of which he is unaware, that has not influenced the data – but the paper doesn't even raise this possibility, let alone eliminate it. Neither does the paper look at the uncertainties involved in the recalibrations employed using proxy data nor at the limitations of the correlations on which they are based.

Having said that, the paper does do enough to suggest that the Helsinki data in the years identified do have sufficient problems to merit a recalibration. Thus the paper is worthwhile and, given there are very few reliable solar-terrestrial data sequences available for these years, it is also quite important to draw attention to the potential problems. So unlike referee 2, I do not recommend rejection. I would recommend publication after a re-write to introduce more rigour and care (and a little more humility) into the wording. Acceptable wording would be to the effect that proxies suggest that the Helsinki data in these years need recalibrating, that the optimum factor is around 30% for the Horizontal component... etc. Definitive statements should not be used unless backed up by statistics – personally, I think they require rejection of the null hypothesis at the 3-sigma level (whereas “probable” statements need backing up at the 2-sigma level and “indications” by the 1-sigma level).

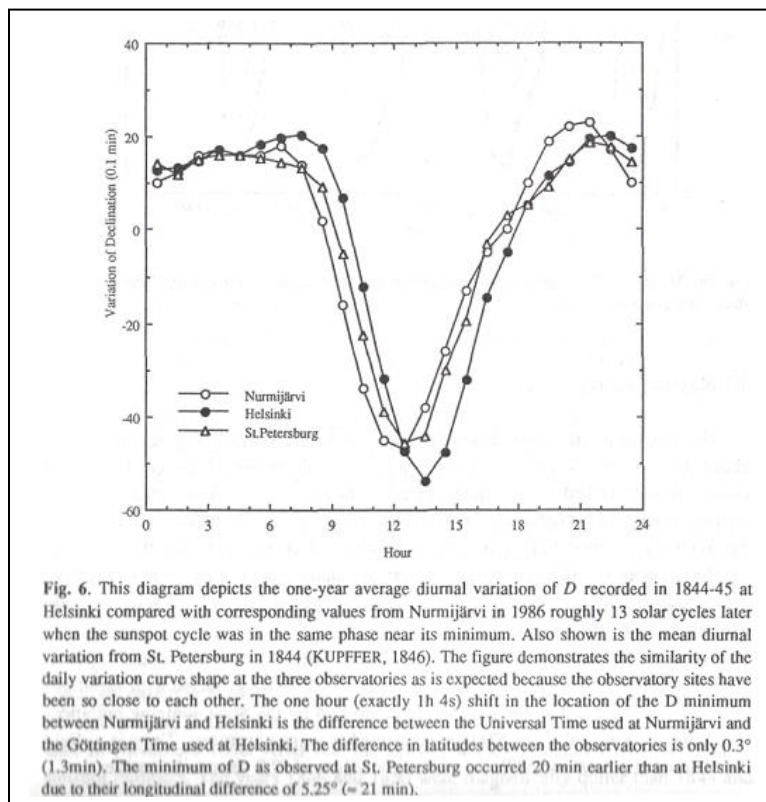
Response to Referee #1:

>>> I note that the author suggests that "metadata" be used to correct the miscalibration, line 180. The author says these metadata should be found and used to make the correction.

<<< I have over the past decade urged Nevanlinna to begin/continue a search for metadata. To date, none has been found as perhaps the issue has not been taken seriously enough. Perhaps publication of the paper might inspire or cajole the authors to get back on the job and do something about it.

>>> There is no alternative, and the author needs to make this clear.

<<< If no metadata can be found, there are two choices: 1) reject the data or 2) compare the data with other stations or with independent data. Rejecting the data is too harsh as there is a large information content that would be lost. But, contrary to the assertion of the referee, there are alternatives: There is a well-established, strong, and physically understood relationship between the range of the diurnal variation and the sunspot number (and the number of sunspot groups). This relationship was discovered by Rudolf Wolf and by Jean-Alfred Gautier as long ago as 1852. Figures 5, 11-13 demonstrate that the Helsinki data are wrongly calibrated for the times indicated and that a single factor [for each of the elements] is enough to bring Helsinki in line with the variation expected from the solar variation and from the data for other stations at that time.



In their original description of the Helsinki data (Geophysical Publications, vol. 27, 1992) Nevanlinna et al. correctly uses that very same relationship [see their Figure at left] to validate the quality of the Helsinki data by direct comparison with other stations (Nurmijarvi and St. Petersburg) at times with similar level of solar activity.

The referee could do well comparing this Figure to my Figures 11 & 12 and to contemplate the obvious educational implications which clearly seem to have escaped him.

>>> Data that have problems cannot be "fixed" without actual auxiliary measurements, such as the calibration metadata discussed by Svalgaard.

<<< Data from other stations (e.g. Colaba, Prague, and Greenwich) constitute 'actual auxiliary measurements'.

>>> Anything else, such as estimates based on inference (like that reported in this manuscript) is not acceptable.

<<< But since auxiliary measurements can be brought to bear, the estimates are acceptable and necessary.

>>> Many things are not known. For example, are there multiple miscalibrations in the data?

<<< No, the comparisons with the auxiliary measurements show that there is only one [for each element], so, contrary to the referee's assertion, this is known, as would be clear from careful reading of the manuscript.

>>> Is the calibration drifting around for some reason?

<<< As the auxiliary measurements show that a single constant factor will do, then contrary to the referee's doubt there is no drift observable in the data. It is, of course, possible that there are drifts below the threshold of observability so this fascinating subject can not be excluded, but also cannot have significant influence.

>>> When, exactly, did the miscalibration begin? When was it fixed?

<<< The paper shows when the miscalibration began and when it was fixed.

>>> How do we know the answers to these questions?

<<< We know the answers to these questions by careful analysis and comparisons with auxiliary measurements taken at the same time at other stations.

>>> A statistical analysis (like that reported here) cannot answer these questions with precision.

<<< Such analysis only needs to answer these questions with a precision sufficient for the reconstruction.

>>> It can only help us understand that a possible problem exists.

<<< Contrary to the referee's assertion such comparative analysis can help us to both detect a problem and to overcome it.

>>> Unfortunately, as the manuscript is presently written, observatory workers might be tempted to "fix" their data archives.

<<< One might sincerely hope that observatory workers would, indeed, pay more attention to the quality of their data archives and to correct problems when identified. The Helsinki data was downloaded from <http://wdc.kugi.kyoto-u.ac.jp/> and cleaned for the usual types of errors found in such datasets, like data for November 31st, time being off by one hour from 1882 on, and base values for Declination being expressed in tens of minutes rather than the (correct) degrees, which causes the Declination to be reported as 54°, about ten times the correct value.

Other examples include <http://www.leif.org/research/MgII%20Calibration.pdf> and the 'mother' of all miscalibrations Figure 5 of <http://www.leif.org/research/IAGA2008LS-final.pdf>

<<< It has happened before, sometimes with unfortunate follow-on consequences.
>>> Perhaps the referee could enlighten me to what unfortunate consequences have followed? And corrections of archival data have, indeed, happened before and to good effect, see e.g. section A4 of <http://www.leif.org/research/2007JA012437.pdf>

<<< So, this issue must be clarified in this manuscript before it is considered acceptable for publication.

>>> As is clear, this issue has been extensively discussed in the manuscript. In fact, the whole paper is precisely aimed at the importance of finding and carefully correcting artifacts in the historical record, based on a thorough understanding of the both the physics and of the observatory practices involved.

The problems identified [and corrected] are not just 'possible problems'; they are blatant and glaring errors and should be strongly marked as such rather than accorded a 'balanced' treatment. The paper is a wake-up call about the vagaries that occur in historical records as well as a demonstration that careful analysis and comparisons can be brought to bear to overcome these difficulties and thus enabling us to put the records to the good use they deserve. I see no reason for not calling a spade a spade and thus no reason to water down the title. I have added 'Correction of' to the title to emphasize the positive aspect of my paper rather than just complaining about the errors.

The proper way for this anonymous referee to voice his concerns is to submit a reasoned Comment Paper with examples of the 'unfortunate consequences' that bother him.

Response to Referee #3:

>>> Acceptable wording would be to the effect that proxies suggest that the Helsinki data in these years need recalibrating, that the optimum factor is around 30% for the Horizontal component... etc.

>>> I will point out that 'proxies' is not the word to use here. I make direct comparison with the same kind of data; see Figures 5, 11-13, not with proxies for the data. As I note in my response to referee #1, that data serves as 'auxiliary measurements' taken at the same time, with the same type of instruments, and responding in a well-understood way to the same stimulus [solar activity]. That the measurements are made at several other stations additionally serves to minimize the possibility of accidental deviations. This should be obvious and hardly require further elaboration.

>>> (S)ometimes based on the knowledge available at that time which turns out to be inadequate, and/or with certain assumptions and/or using limited statistics. The problem is that once a false correction has been made it is often remarkably difficult to track it down and reverse it.

<<< It is of fundamental importance to keep all raw data and to conduct revisions openly and with adequate documentation. With modern digital versions of the data, all adjustments can readily be reversed or improved, so it is not 'remarkably difficult' to undo such corrections. It may be remarkably difficult to overcome resistance to, even justified, corrections, but that is a different, if not more severe, matter. One combats such

‘inertia’ by being blunt about the need for correction and by documenting everything carefully as done in the paper under review.

>>> Definitive statements should not be used unless backed up by statistics – personally, I think they require rejection of the null hypothesis at the 3-sigma level (whereas “probable” statements need backing up at the 2-sigma level and “indications” by the 1-sigma level.

<<< The deviations shown in Figures 11 & 12 are at the 6-sigma level, so I am, indeed, only 99.999,999,803 % sure, which in my book is ‘definitive’ enough.

Now, it seems that it comes down to changing a few words here and there [to show more ‘humility’] and to water down the language to make it more ‘acceptable’. I think this is not appropriate as a purpose of the paper is to needle the authors to be serious about searching for metadata and/or revisit their calibration [now more than 20 years old] before the details and paperwork [notes, computer programs, etc] are lost, in addition to correcting the data archived in the WDC.

Response to Referee #2:

I also want to point to the report of referee #2 which I think should not simply be dismissed, lest you judge the referee to be incompetent at the task. In addition, Lockwood et al. in three papers have already accepted my analysis and made further comparisons and actually revised their reconstruction accordingly. As they write: “a correction to seven years' IDV(1d) data during solar cycle 11 was discussed, checked against newly-analysed independent data from St Petersburg and implemented”. This is a welcome, healthy development, furthering progress and fundamentally dismissing the excessive concerns voiced by referee #1.

Topical Editor Initial Decision: Publish as is (28 Apr 2014) by Prof. Christopher Owen
Comments to the Author:
Dear Dr Svalgaard:

Thank you for your most recent response to the referee reports that I have accumulated regarding this paper. In considering how to proceed, I do indeed note the brief but positive report of referee #2, and the opinion of referee #3 that the paper needed only some minor adjustments before it would be acceptable for publication. Although I note also your overall reluctance to make changes in response to these concerns (perhaps in fact it is your good self who has 'judge(d) the referee(s) to be incompetent at the task?'), I am prepared to allow the 2 positive reports to trump the continued objections of referee #1 and accept the paper for publication. Given the entrenched positions, I am not sure that it is a good use of time to try to achieve a better consensus on this, but I will indeed invite referee #1 to consider voicing his concerns in a comment paper.

Paper accepted